EDITORIAL

The Importance of External Validity

How are research results translated to public health practice? What is the responsibility of researchers, funding agencies, and journals in facilitating the use of research results in public health programs or policies? We address selected aspects of these questions and announce a new emphasis of the *Journal* on external validity for appropriate manuscripts.

IMPORTANCE OF EXTERNAL VALIDITY

Over 40 years ago, Campbell and Stanley published their seminal work on experimental and quasi-experimental designs for research, in which they raised issues about threats to internal validity (whether or not observed covariation should be interpreted as a causal relationship) that exist when researchers are not able to randomly assign participants to treatments.[1] In that volume and subsequent work, they also raised issues about other types of validity, including[2,3]:

(1) Statistical conclusion validity—whether conclusions about statistical inferences of covariation between variables are justified.
(2) Construct validity—whether operational variables adequately represent theoretical constructs.
(3) External validity—whether causal relationships can be generalized to different measures, persons, settings, and times.

It has been frequently argued that internal validity is the priority for research.[4] However, in an applied discipline, the purpose of which includes working to improve the health of the public, it is also important that external validity be emphasized and strengthened.[5–7] For example, it is important to know not only that a program is effective, but that it is likely to be effective in other settings and with other populations.

In an influential 1985 article, "Efficacy and Effectiveness Trials (and Other Phases of Research) in the Development of Health Promotion Programs," Flay proposes a model that emphasizes internal and external validity at different stages of the research process and that would lead to the translation of research to practice.[8] The two main research levels were "efficacy trials" and "effectiveness trials." Efficacy trials were to be highly controlled studies that answered the question of whether a proposed intervention would have the desired effects under ideal circumstances. Effectiveness trials were to follow efficacy trials and were to be studies that carried out the proposed intervention in less controlled and more real-life situations. The argument was that a given public health intervention should be successful in both types of trials before it was ready for dissemination to and by public health practitioners.

Efficacy trials were to have high internal validity, and effectiveness trials were to have high external validity. Efficacy trials were more likely to be controlled experiments, such as randomized control trials of public health interventions, that have the virtue of high internal validity but often have the liability of low external validity[9] (i.e., the groups, settings, or contexts in which findings would apply). It is axiomatic in social science research that there is an inverse relationship between internal and external validity. A key to internal validity is good measurement and study design, and representative sampling is necessary for inference.[9] However, it may be useful to distinguish between inference derived from sample design and our ability to generalize, which is more dependent on judgment.

Historically, researchers have tended to focus on maximizing internal validity, with the idea that it is more important to know if a given public health intervention works under highly controlled conditions than it is to know if it will work among different population groups, organizations, or settings. Similarly, funding organizations and journals have tended to be more concerned with the scientific rigor of intervention studies than with the generalizability of results. The consequence of this emphasis on internal validity has been a lack of attention to and information about external validity, which has contributed to our failure to translate research into public health practice.

For instance, in the area of cancer prevention and control, there is a documented substantial lag between discovery and delivery of effective interventions. Recognition of this lag has been noted for at least 30 years, since the first National Cancer Institute–convened cancer control working groups issued reports in the 1970s. More recently, Balas and Boren found that it takes about 17 years to turn 14% of original research to the benefit of patient care.[10] Similarly, the National Research Council

concluded that, even when effective interventions have been developed, there often is a gap between scientific knowledge and clinical practice.[11] In addition, minorities and underserved communities usually gain access to effective interventions more slowly than do other populations.[12]

Thus, the idea that research would progress from efficacy trials to effectiveness trials to widespread dissemination has not become a reality for a number of reasons, not the least of which is the time and cost involved in this stepwise progress of research to practice.[6] As a result of the failure of this model, practitioners are often unable to determine if a given study's findings apply to their local setting, population staffing, or resources.[6] Reviews indicate that reporting on external validity is provided far less often than is reporting on other methodological issues.[13] However, there are several reasons for the lack of information on external validity being an important contributor to the failure to translate research into public health practice.[14] Policy and administrative decisionmakers are unable to determine the generalizability or breadth of applicability of research findings. Finally, systematic reviews and meta-analyses are limited in the conclusions that can be drawn when external validity data are not reported.

THE *JOURNAL* ENDORSES A GREATER EMPHASIS ON EXTERNAL VALIDITY

Although the *Journal* has long recognized the importance of external validity in articles it has published, the relatively recent CONSORT and TREND reports, as well as the recent emphasis on the RE-AIM model, has strengthened the recognition by the *Journal* editors and editorial board of the need to formally emphasize external validity and to collect information on appropriate manuscripts that enhances both inference and potential generalizability.[6,15–18]

Recently, two members of the the *Journal* editorial board and editors represented the *Journal* in a meeting with 12 other leading health journals and representatives from the National Institutes of Health, the Centers for Disease Control and Prevention, and the Robert Wood Johnson Foundation. The purpose of the meeting was to encourage and strengthen the reporting of findings on external validity. One of the outcomes of the meeting was that participants agreed that enhancing the quality of reporting on external validity in journal articles warrants higher priority than it has received in public health research publications to date.

The meeting participants identified several characteristics of external validity that should be reported. As with other quality-rating scales and guidelines, not every article would be expected to excel on all of the criteria; rather, authors should report on these issues where appropriate, or state that no information is available. Four categories of external validity information were identified by the meeting participants:

1. Study participant recruitment and selection procedures, participation rates, and representative nature at the levels of individuals, intervention staff, and delivery settings.
2. Level and consistency of implementation across program components, settings, staff, and time.
3. Impact on a variety of outcomes, especially those important to populations, practitioners, and decisionmakers (e.g., quality of life, program costs, and adverse consequences).[19]
4. Follow-up reports should include attrition at all levels in item 1, long-term effects on outcomes in item 3, and program sustainability, modification, or discontinuance.

Although we are not intending to add to the burden of authors publishing in the *Journal*, we believe that many of the articles we publish will benefit by including information on external validity. Most important, we believe that the field of public health and public health practice will benefit considerably from this information. ■

Allan Steckler, DrPH
Kenneth R. McLeroy, PhD

About the Authors

Allan Steckler is with the Department of Health Behavior and Health Education, School of Public Health, University of North Carolina, Chapel Hill. Kenneth R. McLeroy is with the Department of Community Health, School of Rural Public Health, Texas A & M University, College Station.

Requests for reprints should be sent to Allan Steckler, DrPH, School of Public Health, University of North Carolina, Chapel Hill, NC 27599-8400 (e-mail: steckler@email.unc.edu).

This editorial was accepted September 20, 2007.

doi:10.2105/AJPH.2007.126847

References

1. Campbell DT, Stanley JC. *Experimental and Quasi Experimental Designs.* Chicago, Ill: Rand McNally; 1966.
2. Cook TD, Campbell DT. The design and conduct of quasi experiments and true experiments in field settings. In: Dunnette MD, ed. *Handbook of Industrial and Organizational Psychology.* Skokie, Ill: Rand McNally; 1976:115–136.
3. Cook TD, Campbell DT. *Quasi-Experimentation.* Chicago, Ill: Rand McNally; 1979.
4. Calder BJ, Phillips LW, Tybout AM. The concept of external validity. *J Consum Res.* 1983;10(1):112–114.
5. Green LW. Evaluation and measurement: some dilemmas for health education. *Am J Public Health.* 1977;67:155–161.
6. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health.* 2003;93:1261–1267.
7. Victora CG, Habicht J, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health.* 2004;94:400–405.
8. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med.* 1986;15:451–474.
9. Bernard HR. *Social Research Methods.* Thousands Oaks, Calif: Sage Publications; 2000.
10. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. In: Bemmel J, McCray AT, eds. *Yearbook of Medical Informatics.* Stuttgart, Germany: Schattauer Publishing; 2000:65–70.
11. Ryff CD and Singer BH, eds.; Committee on Future Directions for Behavioral and Social Science Research at the National Institutes of Health. *New Horizons in Health: An Integrative Approach.* Washington, DC: National Academy Press; 2001.
12. Young WW, Marks SM, Kohler SA, Hsu AY. Dissemination of clinical results: mastectomy versus lumpectomy and radiation therapy. *Med Care.* 1996;34:1003–1017.
13. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? *Ann Behav Med.* 2004;27:3–12.
14. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validity and translation methodology. *Eval Health Prof.* 2006;29:126–153.
15. Turner RJ, Gardner EA, Higgins AC. Epidemiological data for mental health center planning: 1. Field survey methods in social psychiatry: the problem of the lost population. *Am J Public Health.* 1970;60:1040–1051.
16. Luft HS. Regionalization of medical care. *Am J Public Health.* 1985;75:125–126.
17. Moher D, Schulz KF, Altman DG. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* 2001;357:1191–1194.
18. Des Jarlais DC, Lyles C, Crepaz N; the TREND group. Improving the reporting of nonrandomized evaluations of behavioral and public health interventions: the TREND Statement. *Am J Public Health.* 2001;94:361–366.
19. Tunis SR, Stryer DB, Clancey CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290:1624–1632.